Returns to schooling, ability and cognitive skills in Pakistan

Aslam, Monazza; Bari, Faisal; Kingdon, Geeta

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

RECOUP Working Paper No. 20

Returns to Schooling, Ability and Cognitive Skills in Pakistan

Monazza Aslam*, Faisal Bari and Geeta Kingdon

Abstract

This study investigates the economic outcomes of education for wage earners in Pakistan. This is done by analysing the relationship between schooling, cognitive skills and ability on the one hand, and economic activity, occupation, sectoral choice and earnings, on the other. In Pakistan, an important question remains largely unaddressed: what does the coefficient on 'schooling' in conventional earnings function estimates measure? While human capital theory holds that the economic return to an extra year of schooling measures productivity gains acquired through additional schooling, the credentialist view argues that it represents a return to acquired qualifications and credentials while a third, the signalling hypothesis, suggests that is captures a return to native ability. This paper seeks to adjudicate between these theories using data from a unique purpose-designed survey of more than 1000 households in Pakistan, collected in 2007. The paper also examines the shape of the education-earnings relationship in Pakistan as a way of testing the poverty reducing potential of education in Pakistan.

Keywords: rates of return, schooling, ability, cognitive skills, gender, Pakistan
JEL codes: I21, J16, J24.

***Corresponding Author:** Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ, United Kingdom, Telephone: +44-1865-271089.
Email: monazza.aslam@economics.ox.ac.uk

Faisal Bari – Lahore University of Management Sciences, Lahore, Pakistan.
Geeta Kingdon – Institute of Education, University of London, UK.

This paper forms part of the Research Consortium on Educational Outcomes and Poverty RECOUP). Neither DFID nor any of the partner institutions are responsible for any of the views expressed here.For details of the objectives, composition and work of the consortium see: www.educ.cam.ac.uk/RECOUP

Returns to Schooling, Ability and Cognitive Skills in Pakistan

1. Introduction

The objective of this study is to investigate economic outcomes of education for wage earners in Pakistan. This is done by analysing the relationship between schooling, cognitive skills and ability on the one hand, and economic activity, occupation, sectoral choice and earnings, on the other. The labour market benefits of education/skills/ability may accrue both by promoting a person's entry into lucrative occupations and, conditional on occupation, by raising earnings. The aim is to ask whether one or more of these dimensions of human capital raise earnings within any given occupation and/or raise earnings indirectly via facilitating entry into well paying occupations, such as waged work.

Evidence from numerous studies of economic returns to schooling show a consistent pattern of earnings increasing with education. Exactly how schooling generates these economic consequences remains something of a mystery (Katz, 1993)[1]. It is, however, widely accepted that what is learnt is equally, if not more, important than the quantity of schooling acquired (see Glewwe 1996; Ishikawa and Ryan, 2002). Conventional wisdom would suggest that 'years of schooling' are a good proxy for productivity gains in standard estimates of earnings functions or other outcomes. However, this assumption has been challenged in applied work (see Boissiere, Knight and Sabot, 1980, for seminal work on the issue). There is evidence that cognitive skills have economically large effects on individual earnings and on national growth. This literature is summarised in Hanushek (2005)[2] and in Hanushek and Woessmann (2007). A wider body of literature in the US notes the growing importance of cognitive skills as determinants of wages (Murnane, Willet and Levy 1995), that better cognitive skills are associated with greater success in training and on-the-job performance (Bishop 1991), that even high school dropouts' cognitive skills are rewarded (Tyler 2002), and that numeracy skills are a major factor raising the likelihood of full-time employment (Rivera-Batiz, 1992). Currie and Thomas (2001) also note high returns to test scores in terms of wages and employment probabilities among lower socio-economic-status children in Britain. In developing countries the evidence is equally convincing. For instance, Boissiere, Knight and Sabot (1985) find large earnings returns to literacy and numeracy skills of workers in Tanzania and Kenya while Alderman, Behrman, Ross and Sabot (1996) and Behrman, Ross and Sabot (2002) also document statistically significant pay-offs for cognitive skills in the rural wage labour market in Pakistan using data from the 1980s. In a separate

[1] Cited in Ishikawa and Ryan (2002).

[2] Hanushek (2005) cites 3 US studies showing quite consistently that a one standard deviation increase in mathematics test performance at the end of high school in the US translates into 12 per cent higher annual earnings. Hanushek also cites three studies from the UK and Canada showing strong productivity returns to both numeracy and literacy skills. Substantial returns to cognitive skills also hold across the developing countries for which studies have been carried out, i.e. in Ghana, Kenya, Tanzania, Morocco, Pakistan and South Africa. Hanushek and Zhang (2006) confirm significant economic returns to literacy for 13 countries on which literacy data were available.

vein, Jolliffe (1998) finds that cognitive skills have a positive effect on total and off-farm income of Ghanaian households.

The key problem lies in identifying what exactly the coefficient on 'schooling' in conventional earnings functions measures. While human capital theory advocates that the economic return to an extra year of schooling measures productivity gains acquired through additional schooling, the credentialist view argues that it represents a return to acquired qualifications and credentials while a third, (the signalling hypothesis) suggests that it captures a return to native ability. Paucity of data has meant that it has not been possible to ask which of these alternative interpretations applies in Pakistan, and more generally elsewhere[3]. This study aims to address this question both within the context of occupational attainment and earnings functions. It asks: is it productivity gains, credentialism or native ability that raise earnings by facilitating entry into more rewarding occupations, and which of these directly raise individual earnings *within* waged work?

An answer to this question is important in identifying the value (if any) of acquiring additional education, and has implications for education policy. For instance, if education merely serves as a signal for higher ability but does not of itself raise worker productivity, then the efficiency rationale for public funding of education disappears, even though schooling remains privately profitable as it enables individuals to command higher wages through signalling ability. Addressing these concerns is a main objectives of this study.

The paper also examines the shape of the education-earnings relationship in Pakistan. While it has conventionally been assumed that earnings increase with education at a decreasing rate (i.e. the relationship is concave), data for Pakistan for the late 1990s suggest that the relationship is convex, i.e. that the return to an extra year of education progressively increases with education level. We re-examine this issue with very recent data to see whether the extent of convexity has been maintained or increased over time. The data come from a unique purpose-designed survey of more than 1000 households. The data were collected in 2006-2007 from nine districts in Punjab and the North West Frontier Province (NWFP) and of Pakistan. As well as containing standard information needed for the estimation of earnings functions (years of schooling, work experience and earnings), the data also uniquely include measures of individuals' cognitive skills (scores on tests of literacy and numeracy) and of their innate ability (score on Raven's Progressive Matrices test). We use a sample of urban and rural wage earners aged 15-60 and estimate occupational and sectoral attainment functions and earnings functions separately for males and females to allow for the vector of coefficients to differ across gender. Occupational and sectoral attainment models are estimated using multinomial logits (and binary logit in one instance) while earnings functions are estimated using Ordinary Least Squares

[3] While Alderman et al.'s (1996) study addresses this question for wage earners in Pakistan, it does so only for males in rural areas and the findings are based on data from the 1980s. The authors find large positive returns to cognitive skills and no return either to years of schooling or ability. The authors attribute this finding to suggest that the rural labour market in Pakistan rewards human capital.

(OLS), Instrumental Variables (IV) and household fixed-effects (see Section 2 for details). The models include (among the standard vector of variables) measures of years of completed schooling, cognitive skills and innate ability. The cognitive skills variables used are further split into 'basic-order skills' – a dummy variable measuring whether a person is literate or not – and 'higher order skills' measuring a larger set of literacy skills (more details in Section 3).

There are several interesting findings. Firstly, the evidence shows that while basic-order skills promote women's entry into lucrative wage occupations (and within them into blue-collar jobs), for men a wider cognitive skills-set is required to aid entry into the more rewarding occupations. Much of the direct effect of cognitive skills disappears after conditioning on schooling suggesting that the effect of cognitive skills operates through schooling attainment. Secondly, there is no direct effect of ability in the earnings functions either for men or women. That OLS, IV and household fixed-effect estimates also consistently reveal an economic rate of return of about 5 per cent for all wage earners, confirm the lack of 'ability bias' and there appears to be no evidence of signalling. Thirdly, the estimates reflect a direct return to schooling for men (and no return to cognitive skills) while for women there is a suggestion of a return to cognitive skills (in the OLS estimates). The data support the human capital hypothesis for women and a credentialist hypothesis for men. Finally, while we can attribute the return to schooling for men to credentialism, further investigation shows that much of the effect of schooling operates through positive behavioural traits possessed by individuals when aged 15. Thus, a direct return to schooling may not simply reflect credentialism and could be seen to reflect a return to non-cognitive traits valued (and hence remunerated) in the labour market. Evidence also confirms the profound convexity of education-earnings profiles observed in previous studies in Pakistan.

This paper proceeds as follows. Section 2 discusses the underlying methodology while Section 3 describes the data. Section 4 reports the main results and Section 5 concludes.

2. Methodology

Controls for completed years of schooling, cognitive skills and innate ability are included in both multinomial logit models of occupational attainment and in earnings functions as a means of adjudicating whether the data from Pakistan support credentialism, screening or the human capital hypothesis. The analysis begins by estimating multinomial logit models of occupational attainment. All individuals in the labor market are classified into one of 5 occupational categories: out of the labour force (OLF), unemployed, self-employed in agriculture, non-agriculture self-employed and wage-employees. Unemployed individuals are those who seek employment and are available for it, while OLF individuals are those who do not seek employment, such as housewives, students and the retired.

The second part of the study extends the standard Mincerian earnings function to discern what the coefficient on schooling measures by incorporating measures of cognitive skills and native ability as follows:

$$\text{LnY}_i = \beta_0 + \beta_1 S_i + \beta_2 H_i + \beta_3 A_i + \beta_4 \mathbf{X}_i + \varepsilon_i \qquad (1)$$

In (1), LnY_i is the natural log of earnings[4] of individual i, S_i measures years of completed schooling, H_i and A_i measure cognitive skills and inherent ability respectively, $\mathbf{X}_i$ is a vector of observed characteristics of individual i (such as experience and its square) and ε_i is an individual-specific error term. In this specification, β_1 reflects a return to credentials or non-cognitive traits such as attitudes and interests, β_2 measures the return to productivity or cognitive skills, while β_3 potentially captures the return to native ability. Earnings functions are fitted only on a sample of male and female wage earners aged 15-60.

We start by estimating OLS models of earnings functions to provide some baseline results. OLS estimates potentially suffer from two major biases - sample selectivity and endogeneity (omitted variable) bias. Sample selectivity bias arises due to estimating the earnings function on separate sub-samples of workers, each of which may not be a random draw from the population. This violates a fundamental assumption of the least squares regression model. While modelling occupational outcomes can be a useful exercise in its own right – suggesting the way in which education, skills and ability influence people's decisions to participate in different types of employment – it is also needed for the consistent estimation of earnings functions. Modelling participation in different occupations is the first step of the Heckman procedure to correct for sample selectivity: probabilities predicted by the occupational choice model are used to derive the selectivity term that is used in the earnings function. Following Heckman (1979) and Lee (1983), the earnings equations can be corrected for selectivity by including the inverse of Mills ratio λ_i as an additional explanatory variable in the wage equation.

However, the problem of endogenous sample selection is akin to the problem of endogeneity bias. Endogeneity bias arises if workers' unobserved traits, which are in the error term, are systematically correlated both with included independent variables and with the dependent variable (earnings). For instance, if worker ability is positively correlated with both education/cognitive skills and with earnings then any positive coefficient on education (or cognitive skills) in the earnings function may simply reflect the cross-section correlation between ability on the one hand and both education/skills and earnings on the other, rather than representing a causal effect from education/skills onto earnings.

One approach to addressing endogeneity bias arising from omitted 'ability' is to include variables that proxy for ability directly in earnings functions. Several studies adopt this approach by including measures of 'native ability' scores from tests (such as the Ravens test) in earnings functions. Two of the most notable studies are Griliches (1977) and Boissiere, Knight and Sabot (1985). The only study

[4] We use wage rates as they are a better measure of labour productivity since self-employment earnings incorporate a return to physical capital.

using this procedure in Pakistan we are aware of is by Alderman, Behrman, Ross and Sabot (1996). The consensus from these studies is that the direct effect of ability on earnings is small i.e. 'ability bias' is either small or non-existent. The key criticism of this approach, however, is that test-based ability estimates do not successfully measure individuals' inborn ability and are contaminated by schooling and the environment.

The instrumental Variable (IV) approach provides an alternative solution to the problem of endogeneity. This methodology identifies variables (instruments, W_i) that are correlated with schooling and cognitive skills (S_i and H_i) and uncorrelated with unobserved ability and measurement errors. The key challenge, however, lies in finding suitable instruments i.e. W_i which are not part of the vector **X** in equation 1. Social and natural experiments are useful and many studies use 'institutional variations' in schooling due to such factors as proximity to schools, minimum school-leaving age etc. to instrument for schooling. Card (1995, 1999 and 2001) provides a summary of some of the recent studies that use this approach. The consensus from contemporary research on developed countries is that IV estimates based on natural experiments are as high as and sometimes almost 20 percent higher than corresponding OLS estimates (Card, 2001). The evidence from developing countries is similar (see Maluccio, 1998 and Duflo, 2001).

However, natural-experiment-based IV approaches have exacting data demands and an alternative is to use non-experimental IVs for endogenous schooling. As children's schooling outcomes are to a large extent driven by family background (FB), variables such as father's and mother's education or 'distance to school' are sometimes used (Söderbom *et al.*, 2005 and Trostel *et al.*, 2002 are examples of two recent studies using parental education as instruments for schooling). FB variables constitute valid instruments if they affect earnings only indirectly through their effect on schooling or cognitive skills, i.e. if there is no intergenerational transmission of ability. With this caveat, we use family background variables as instruments for schooling and cognitive skills. FB then enters the vector of variables which directly influence schooling and cognitive skills[5].

An alternative to the IV technique is to use observations from different individuals within the same family to estimate a 'household fixed effects' earnings equation. To the extent that unobserved traits are shared within the family, their effect will be netted out in a family differenced model. For instance, the error term 'difference in ability between members' will be zero if it is the case that ability is equal among members. While it is unlikely to be the case that unobserved traits are identical across family members, it is likely that they are much more similar within a family than across families and, as such, family fixed effects estimation reduces endogeneity bias without necessarily eliminating it entirely.

In this study, we adopt all three approaches – the direct 'ability' proxy, IV and household fixed-effects – in a bid to discern the causal impact of skills and schooling on earnings.

[5] Alderman et al. (1996) use several school quality indicators as well as parental education to instrument for endogenous cognitive skills.

3. Data and Descriptive Statistics

The data for this study come from a purpose-designed household survey administered to 1194 urban and rural households between November 2006 and March 2007. Households were selected randomly through stratified sampling from 9 districts in two provinces – Punjab and the North West Frontier Province (NWFP) - in Pakistan[6]. The data were collected under the auspices of the Research Consortium on Educational Outcomes and Poverty (RECOUP).

The survey collected rich information on various individual, family and community-level factors. While the roster captured basic demographic, anthropometric, education and labour market status information on *all* resident household members in the sampled households (more than 8000 individuals), detailed individual-level questionnaires were administered only to those aged between 15 and 60 years. Some 4907 individual-level questionnaires were filled. These individuals were also administered tests of literacy, numeracy, health knowledge, English language and the Ravens Progressive Matrices test (to assess innate ability). The first three of these – literacy, numeracy and the health knowledge test – were translated into Urdu, the national language. The literacy and numeracy instruments were designed to capture 'basic order' skills and 'higher order' skills. For example, the first half of the literacy test consisted of a small passage followed by a few questions testing reading comprehension. Only if a person could answer three out of the total of five questions correctly in the short test was he/she administered the 'long literacy test' which tested more advanced reading and comprehension skills. The numeracy test was also designed similarly. Since the objective of the Mathematics test was to measure numeracy skills rather than ability to read, the enumerators were trained to read out the numeracy test questions to the respondents. While both the short and long literacy tests appear to have captured 'basic order' and 'higher order' literacy well (see Figure 1), the short numeracy test unfortunately turned out not to be a good measure of mathematics skills for the sample of wage earners. As we are not convinced of the validity of the short numeracy test, we do not use the scores from the 'basic order' numeracy skills test in any of the regression analyses. The total mathematics score (sum of the short and long numeracy tests), measuring 'higher order skills' is used in some regressions discussed later. The analysis in this paper is restricted to the sub-sample of waged workers aged 15-60 for whom we have literacy and numeracy test scores available.

Test scores non-response was greater for men than for women largely because they were less likely to be available at home for the tests at the time of the enumerator's visit during the survey. While this is likely to generate sample selectivity concerns if individuals for whom scores are available are non-randomly drawn from the population, data limitations prevent us from correcting for any biases that may consequently arise.

[6] Rahimyar Khan, Khanewal, Sargodha, Kasur, Attock and Chakwal districts were chosen from Punjab while Swaat, Charsadda and Haripur were sampled from NWFP.

Table 1 shows summary statistics for labour market status by gender. The extent of gender asymmetry in the distribution of the labour force is striking. A significant majority of women are not economically active - about 70 per cent women are out of the labour force compared to 18 per cent men. Among men who participate in the labour force, wage employment absorbs the highest proportion (42 per cent). This is true only for 7.5 per cent of women. This implies that our sample of wage earners, especially women, is unlikely to be randomly drawn from the adult population. Table 2 confirms this suspicion – a much larger proportion of waged women (31 per cent) than men has completed 12 or more years of education. Given that among *all* women (aged 15-60), only about 9 per cent have completed 12 or more years of education, the sample of wage employees is indeed atypical. Thus, women entering waged work have either very low levels of education and enter waged work due to necessity or have completed 10 years of education or more (50 per cent) and enter waged work because they aspire to do so.

[Table 1 about here]

[Table 2 about here]

[Table 3 about here]

Table 3 summarises some key statistics. Column (a) shows some interesting differences in mean years of completed schooling for different labour market categories. While many of these differences are unsurprising, several are striking. For instance, there is more than a 2 year difference in mean years of schooling between rural and urban regions. The largest gap (of almost 8 years of schooling) is between white-collar and elementary workers. The second largest gap is between government and private school graduates. Columns (b) and (c) show the proportions of individuals who are literate or numerate (LITERATE was defined as a dummy variable equalling 1 if an individual scored 1 or more, i.e. up to 5 in the short literacy test and 0 otherwise[7]. A person is defined to be numerate if they scored at least 3 out of 5 in the short mathematics test and 0 otherwise) in the various categories. There appears to be little variation in NUMERATE which suggests that, at least among the sample of wage earners, the short numeracy test was not able to distinguish between numerate and non-numerate individuals. Column (e) reports mean earnings (Rupees. per month). Consistent with evidence elsewhere, the earnings gap between males and females is large and significant. While we expected larger differences in wage earnings of rural and urban employees, the insignificantly small difference can be explained by the relatively large proportion (25 per cent) of individuals employed in the government sector in rural regions.

[7] While a person was defined to be LITERATE and, hence, assumed to possess 'basic order literacy skills', if they scored 1 or more in the short literacy test, simple cross-tabulations revealed that for 35 individuals (33 men and 2 women) LITERATE equalled zero (i.e. they were deemed illiterate) despite having completed 5 years of education or more. Further investigation revealed that for all 35 individuals a language other than Urdu was spoken at home. Our tests, however, were administered in Urdu. Therefore, these people's inability to comprehend the literacy test possibly captured their inability to comprehend the passage in Urdu rather than reflecting their literacy status. Hence, for these 35 persons, LITERATE was replaced to equal 1.

4. Empirical Findings

4.1 Occupational Attainment

This sub-section investigates whether one of the ways in which the labour market benefits of human capital accrue is via promoting entry into more lucrative occupations and deterring entry into less-rewarding ones. While a recent paper by Kingdon and Söderbom (2007) investigates this question for Pakistan, our paper differs from theirs in a number of ways. Unlike their study which uses data from 1999, ours uses latest data from 2007. Also, while their dataset has the advantage of being representative of all provinces from Pakistan (ours is from the Punjab and NWFP), the richness of the RECOUP dataset allows us to go beyond what they have achieved. For instance, Kingdon and Söderbom's measures of cognitive skills are based on self reports of 'whether an individual can read or write in any language' (equals 1 if literate) and 'whether person can do simple sums' (equals 1 if numerate). Our rich data allow us to look not just at the impact of education and cognitive skills but also of innate ability. Moreover, our cognitive skills measures are based on actual tests of literacy and numeracy rather than on self-reports[8]. Also, the instruments were devised as short and long tests so as to capture basic ability to read and do maths ('basic order skills') and advanced reading comprehension and mathematics skills ('higher order skills'). Finally, our rich data allow us to investigate the impact of the various dimensions of human capital not just on broad occupational outcome (out of the labour force (OLF), unemployed, agriculturally self employed, non-agriculturally self employed and waged workers) but also allows division of waged workers into those in the government or private sectors and those employed either in elementary occupations, blue-collar work or white-collar jobs.

Table 4A reports marginal effects from multinomial logit models estimated on activity status (in the first category) controlling for basic-order literacy (LITERATE) and the ravens score (among the standard controls reported in the table). Table 4B additionally controls for schooling (in quadratics). Focus first on the findings in Table 4A which show the direct effects of literacy and innate ability without conditioning on years of schooling[9]. The estimated vector of coefficients varies greatly by gender. While being literate clearly promotes women's entry into the more lucrative wage employment sector and reduces the likelihood of being in less-rewarding agricultural self-employment, possessing basic order skills reduces men's likelihood of being in wage or agricultural employment and increases chances of being OLF or unemployed[10]. Marginal effects on the ravens

[8] Figure 2 illustrates that while cognitive skills are clearly correlated with schooling – individuals perform better on these tests with higher levels of completed schooling – the ravens score is fairly invariant to schooling. The latter finding gives some credence to the requirement that our measure of 'ability' does not depend entirely on schooling.

[9] We show the direct effect of literacy/ability without conditioning on schooling because these dimensions of human capital are highly correlated.

[10] It seems surprising that those who possess basic literacy (which includes individuals with higher order literacy skills) have higher chances of being OLF. But the fact that possession of higher order literacy skills does not increase chances of being OLF (Table 5A) suggests that the positive relationship between basic literacy and being OLF in Table 4A is due to the greater influence of individuals who have only the basic literacy skills.

score suggest that more able men (i.e. with higher scores on the test) are more likely to be in non-agricultural self employment and less likely to be in agriculture. Among women, the more able are less likely to be out of the labour force and in waged work and also more likely to be unemployed (possibly because they willingly wait for better job opportunities).

[Table 4A about here]

As suspected, conditioning on schooling in Table 4B causes much of the direct effect of LITERATE to disappear (except by continuing to increase men's likelihood of being out of the labour force and reducing both men and women's chances of employment in the less well-paying agricultural sector). Among men, there is a convex relationship between years of schooling and the probability of being in wage employment: as also seen clearly in Figure 3, the chances of being wage employed first fall with education and then rise with education beyond 6 years of schooling. Thus, there is a threshold effect: low levels of schooling do not permit men to join wage employment but beyond 6 years, their probability of joining waged work increases rapidly at an increasing rate. The effect of schooling increases at a decreasing rate for men in non-agricultural self employment and the turning point occurs at about 6.6 years of education. Among women, greater schooling increases the chances of being in wage employment (the coefficient on schooling is almost significant at the 10 per cent level). Importantly, it is only after women have acquired 7 years of schooling or more that their probability of being out of the labour force sharply declines. Given that only 29 per cent women have completed 7 years of education or more suggests that a relatively small proportion of women can accrue the labour market benefits of schooling. These findings are graphically depicted in Figures 3 and 4.

[Table 4B about here]

[Figure 3 about here]

[Figure 4 about here]

The aforementioned findings are not inconsistent with those reported in Kingdon and Söderbom (2007) and in Aslam, Kingdon and Söderbom (2007). The key difference in this study, however, is the availability of what we term 'higher order skills' i.e. individual's total scores in literacy and mathematics which were not available in these recent studies. Occupational attainment may depend not just on whether the individual is literate but also on how strong literacy and numeracy skills are. Tables 5A and 5B report the marginal effects from multinomial logit models conditioning on higher order skills and the ravens score and on these and schooling, respectively.

[Table 5A about here]

Focus first on table 5A. There are several interesting findings. First, higher order literacy and numeracy skills clearly improve men's chances of being in lucrative waged work. The positive effects of higher literacy and numeracy skills operate after men score about 3.7 points on the literacy and 4.8 points on the numeracy tests. This is true for about 46 per cent and 59 per cent men in the sample. For women, on the other hand, the direct effect of higher order skills only just works in reducing the likelihood of being unemployed (achieving 5.6 points or more in the mathematics test reduces likelihood of being unemployed). However, as only 15 per cent women achieve 6 points or more in the numeracy tests, this positive effect affects a relatively small proportion of women.

Table 5B conditions on schooling (in quadratics). The direct effect of higher order skills disappears after conditioning on schooling for men though for women, we note that for any given number of years of education, the higher the literacy skills, the greater the chances of being in non-agricultural self-employment.

[Table 5B about here]

Summarising, the findings from Tables 4 and 5 suggest that while possessing basic order skills improves women's chances of being wage employed, men's labour market choices are determined more by a fuller set of cognitive skills. Part of the explanation for this finding is the relative scarcity of well-skilled women compared to men. Alternatively, it could be that *within* these broadly defined labour market categories, the types of occupations and jobs done by men and women differ in the requirement of skills – i.e. women are employed in jobs that require possessing mostly basic skills while men's occupational choices require them to hold a superior skills set. For instance, among 'literate' women, a majority (86 per cent) are concentrated in the following two categories: life science and health associate professionals (midwives and nurses etc.) and teaching associate professionals (primary or pre-primary education). Just 'literate' men, on the other hand, are fairly evenly distributed across the occupational spectrum.

The impact of 'literacy' on occupational and sectoral choice among the wage employed is partly investigated in Tables 6 and 7. Table 6 reports the marginal effects from a logit estimate of whether a waged worker is employed in the government sector (equals 1) or the private sector (equals 0). Years of schooling and the possession of higher order numeracy skills cause men to be more likely to get government as opposed to private sector jobs. Table 7 reports marginal effects from a multinomial logit on occupational choices (elementary, blue-collar and white-collar jobs) estimated for waged workers. 'Elementary' corresponds to unskilled, 'blue collar' to semi-skilled and 'white-collar' to skilled jobs. Clearly, among this classification of waged workers, maths skills promote women's entry into blue-collar jobs. Surprisingly, neither schooling nor cognitive skills appear to affect waged men's occupational attainment as between elementary, blue-collar and white sector jobs. However, when the extent of male joblessness is as high as it is in Pakistan – according to Table 1 nearly 23% of men are either unemployed or OLF (which is disguised unemployment among prime age men) – then perhaps even educated men with good cognitive skills cannot be choosy as between the type of waged job they will accept.

[Table 6 about here]

[Table 7 about here]

4.2 Earnings Functions

This sub-section discusses the findings from estimates of semi-logarithmic earnings functions. Table 8 showcases standard Mincerian earnings equations controlling only for years of completed schooling and labour market experience for the pooled sample (men and women) and separately for the two genders. The first column shows the linear while the second the quadratic specification respectively. The large unconditional earnings premium observed for men in the raw descriptive statistics in Table 3 is reflected in an even larger conditional premium in Table 8. While the economic rate of return to schooling (for each additional year acquired) is about 5 per cent for both men and women, consistent with past evidence from Pakistan, the economic return to women's schooling is almost double that for men - 8.3 per cent for the former compared to 4.5 per cent for the latter (see Aslam, 2007, Kingdon and Söderbom, 2007 and Riboud, Savchenko, and Tan, 2006 for some recent studies in Pakistan reporting higher labour market returns to women's schooling than men's). Finally, in this basic specification, while the relationship between schooling and earnings is quadratic for men, it appears to be linear for the sub-sample of women. Interestingly, inclusion of the ravens test score in the specifications reported above doesn't alter any of the aforementioned results, unsurprising given the insignificance of innate ability in the earnings functions specifications (see Appendix A1). This finding suggests that the coefficient on schooling doesn't appear to capture a return to native ability. Parsimonious specifications from this point onwards do not include the ravens score in earnings function estimates.

[Table 8 about here]

Table 9 reports earnings functions estimates incorporating a measure of basic-order literacy skills (as mentioned before, we do not include any measures for basic-order numeracy skills). LITERATE is insignificant in both the linear and quadratic specifications for men. For women, it is positive and precisely determined in the quadratic specification.

A comparison between column (3) in Table 8 and column (3) in Table 9 shows that for men, the coefficient on schooling remains significant and doesn't change substantially when direct measures of cognitive skills are introduced. This holds for the quadratic specification as well. A comparison for females (column 5 in Table 8 with column 3 in Table 9) reveals that the inclusion of cognitive skills measures causes the coefficient on schooling to fall to zero (from 0.083) suggesting that schooling contributes for women directly through cognitive skills acquisition though the effect is small and not significant. This finding does change in the quadratic specification – both coefficients on schooling and LITERATE are significant signifying that both contribute towards women's earnings[11]. There are two possible interpretations of the result for men: firstly that schooling attainment directly makes a

[11] Alternatively, earnings functions were estimated with basic-order literacy skills measured in the form of 5 dummy variables – DUMLIT1, DUMLIT2, DUMLIT3, DUMLIT4 and DUMLIT5. DUMLIT1 equals 1 if an individual scored 1 in the short literacy test and 0 otherwise, DUMLIT2 equals 1 if an individual scored 2 in the short literacy test etc. The objective was to determine whether the effect of possessing better basic-order literacy skills was any different from lesser ability to read (measured through a smaller score). The findings are reported in Appendix A2. Clearly, the coefficients on the 5 literacy dummy variables are mostly not significantly different from each other implying that one can collapse basic-order literacy into LITERATE.

contribution to earnings over and above any arising from cognitive skills acquisition (i.e. there is an element of credentialism for waged men), and secondly that the cognitive skills variables does not adequately measure cognitive skills differences between leavers at each education level so that schooling attainment itself is a better proxy for cognitive skills than the LITERATE variable. Which interpretation is more plausible? Since cognitive achievement increases with schooling (Figure 2) and LITERATE is a dichotomous variable that is unlikely to capture all aspects of literacy that result from schooling, we are somewhat hesitant to conclude that there is credentialism for waged men in Pakistan. This reluctance is further strengthened when we consider that cognitive skills scores are available only on a sub-sample of waged men, resulting in greater selectivity bias than in the case of women.

[Table 9 about here]

While basic-order cognitive skills captured in the short literacy tests are rewarded in the form of higher earnings for women, possession of basic-order literacy skills does not guarantee a return for men. Part of the explanation for this differential reflects a scarcity premium – far fewer women are literate according to our definition compared to men (59 per cent women compared to about 69 per cent men). Our findings also indicate a direct return to schooling for men (linear and quadratic specifications) and women (only the quadratic specification). This could reflect credentialism or screening or even human capital acquired in school or at home. Moreover, it could even represent a return to non-cognitive traits such as social and interpersonal skills and attitudes which may be acquired through schooling and valued (and hence remunerated) in the labour market.

Table 10 extends the analysis by estimating earnings functions with basic-order skills by region. For men, there are direct returns to schooling in both rural and urban regions (columns 1 and 2). For women, there are large positive returns to schooling in urban areas and no returns in rural regions. Importantly, as before, men's cognitive skills are not directly rewarded in the form of higher earnings. While literate women are positively remunerated in rural areas, the coefficient on LITERATE is perversely negative in urban regions (though it is not significant in the quadratic specification).

[Table 10 about here]

Summarising, the story so far reveals that for men the labour market appears to reward schooling attainment. For women, the rewards operate relatively more through basic-order cognitive skills rather than through years of completed schooling. Appendix tables A3-A5 extend this work by estimating earnings functions incorporating higher order skills (A3) and separately for different worker groups and categories in linear and quadratic specifications (A4 and A5 respectively). The findings emerging from these tables are that higher-order cognitive skills are not rewarded for either gender and that while basic-order skills are rewarded for different classifications of women-workers, the results are sensitive to whether schooling is specified as linear or quadratic.

4.3 Robustness Checks

4.3.1 Socialisation Skills and Omitted Variable Bias

While the cognitive skills tests used may be quantifying academic learning (albeit imperfectly), the labour market may reward non-academic 'socialisation skills' (such as discipline, networking, planning and leadership ability) equally if not more than academic expertise. These social skills may not be fully captured in measures of academic performance. Individuals who attain higher education levels, score better in standardised tests and earn more, may be those who at a young age have certain personality traits conducive to learning and success in other outcomes. These skills may or may not be learnt while at school. For instance, a study by Cawley, Heckman and Vytlacil (2001) in the US finds that 'socialisation skills' are important determinants of labour market outcomes along with cognitive skills.

The RECOUP dataset includes seven measures of self-reported 'personality traits' for individuals aged 15-60. Individuals were asked to rank using a rating scale ranging from 1 (very inaccurate) to 5 (very accurate), how they believed their personality type was (when aged 15) on the following dimensions: making friends easily, finding it difficult to get down to work, having a lot to say, making plans and sticking to them, accepting people as they were, being sympathetic to religious parties and causes and doing just enough work to get by. We converted these variables into zero-one indicator variables and generated an equally weighted index of 'positive personality traits' (making friends easily, had a lot to say, made plans and stuck to them and accepted people as they were) and an index of 'negative personality traits' (found it difficult to get to work and did just enough work to get by). As individual's sympathy to religious or other causes could be deemed as a positive or negative trait depending on the circumstance, RELIGION (equals 1 if were sympathetic and 0 otherwise) was kept as a separate indicator variable and not included in either of the two indices[12].

In Table 11 we examine the relationship between behavioural and personality traits when aged 15 and log earnings subsequently for all wage earners. In particular, we regress log earnings on personality traits without both schooling attainment and cognitive skills and then with each of these and finally condition on both. When we do not condition on schooling and cognitive skills, we find positive socialisation/behavioural effects and favourable religious attitudes have a positive effect on log earning. For instance, individuals reporting positive personality traits when aged 15 find themselves earning up to 6 per cent more than those without such reported characteristics. However, these effects disappear completely when conditioning on schooling attainment or basic-order cognitive skills (separately or together) suggesting that schooling and positive socialisation skills are highly collinear. Thus, rather than having a direct effect on earnings, much of the effect of personality traits operates through educational attainment (rather than through cognitive skills) and these findings

[12] The proportions reporting '1' among the positive traits were: friends (67%), had lot to say (47%), stuck to plans (52%) and accepted people (65%). Proportions reporting '1' in negative traits were: difficult to work (52%) and did just enough work (65%). Proportion reporting sympathy for religious/other causes was 42%.

imply that while employers reward positive socialisation skills in the form of higher earnings, they view these skills as having developed through schooling and the return to higher schooling attainment is partly a reward for possessing favourable traits. This suggests that our finding of a direct return to schooling may not simply reflect credentialism but could be reflecting a return to non-cognitive traits valued (and hence remunerated) in the labour market.

[Table 11 about here]

4.3.2 Addressing Endogeneity of Schooling and Cognitive Skills

The discussion so far has assumed that the coefficients on both schooling and cognitive skills are measured consistently and without bias i.e. they measure the causal impact of both on earnings. As mentioned in Section 2, one of the greatest challenges facing researchers is the difficulty of attributing the impact of either as causal when faced with unobserved traits (the classic 'ability bias' problem). The coefficients on schooling and cognitive skills can only be interpreted as causal effects if earnings differentials between individuals with varying years of schooling/skills do not reflect differences in ability or in other unobserved investments in human capital that happen to be correlated with education[13].

As mentioned before, including proxies for ability directly in earnings functions is one way of approaching this problem. Much work adopting this approach finds that including a direct measure of 'ability' in earnings functions does not substantially or significantly reduce the coefficient on schooling, leading to the conclusion that the direct contribution of 'ability' on earnings is small (Griliches, 1977). Indeed this is the finding not only from past work on Tanzania and Kenya (Boissiere, Knight and Sabot 1985) and Pakistan (Alderman *et al.* 1996) but also from the current study when a measure of the Ravens score is included in earnings functions (see Appendix Table A1). Unfortunately, very few datasets contain ability measures that are convincing and, in the present study, it cannot be guaranteed that the Ravens score is not reflecting the contribution of schooling or the environment.

Alternative approaches to addressing the endogeneity problem include using Instrumental Variables (IV) or household fixed-effects approaches (discussed in Section 2). The former is based on

[13] Due to space constraints, we relegate the selectivity-corrected results (with inverse mills ratios computed from the multinomial logits of occupational choice reported in Tables 4A and 4B) to this footnote. We replicated Table 9 controlling for selection into waged work and estimated 8 regressions (linear and quadratic specifications of the earnings functions for males and females with lambda terms computed from multinomial logit models based on specification 1 in Table 4A and specification 2 in Table 4B). In all 8 regressions, lambda is negative but significant only in one instance- for males with schooling specified as linear and lambda computed from Table 4B (of magnitude -0.518). These findings largely suggest that sample selection is not a major concern at least in the data on which this study is based. Two recent studies from Pakistan (Aslam 2007 and Kingdon and Söderbom 2007) report an upward bias in the baseline OLS estimates of earnings functions on wage earners arising from selectivity. This could partly be explained by their larger sample sizes and national representativeness (both studies are based on the 1999 and 2001 waves of the nationally representative Pakistan Integrated Household Survey covering more than 14000 households from all provinces and regions in Pakistan).

finding credible instruments correlated with schooling and cognitive skills but not correlated with the residual in the earnings function. The latter approach uses observations from different individuals within the same family to 'difference out' the variables generating correlation in the residuals (akin to the twin-differencing approach).

Tables 12A and 12B report OLS and IV earnings functions estimates for all persons and separately for males and females. The vector of instruments used to control for the endogeneity of schooling attainment and of cognitive skills includes: parental education (fathers and mother's completed years of schooling), distance to primary school in meters (when individual was of primary school-going age) and its square, and the individual's ravens test score. This instruments set appear to be 'valid'. The first stage estimates reported in column (3) in Tables 12A and 12B show that in almost all instances, the instruments have large, precisely determined coefficients with expected signs. The instruments are also accepted as valid and in no instance does the p-value of Hansen's over-identification test reject the null hypothesis at the 5% level of significance.

[Table 12A about here]
[Table 12 B about here]

Focus on the findings in Table 12B. For men, there are direct returns to schooling but no positive returns to literacy. The marginal rate of return to education (RORE) is about 10%
using IV. The RORE also doubles when using IV compared to the OLS on the sample on which IV was estimated. This finding is consistent with evidence from previous studies and also recently in Pakistan (see Aslam 2007)[14]. The RORE for women also increases when estimated using IV (from about 6% to about 7%), but is statistically insignificant in both OLS and IV equations, possibly due to the small sample size. There also appears to be no return to literacy for women using either approach. While part of the positive coefficient on schooling for men may be attributed to credentialism, Table 11 above has shown that some of it is a return to non-cognitive skills (behavioural and personality traits) rewarded by employers in the labour market.

We turn next to household fixed-effects estimates of returns to education. The results are based on the sub-sample of households that had at least one male and one female wage-worker who are related in any way (father-daughter, mother-son, brother-sister, husband-wife etc.)[15]. While this

[14] One explanation for this was proposed by Griliches (1977) who suggested that upward ability bias in the OLS estimator is small so that the difference between the OLS and IV estimators reflects correction for attenuation bias arising from measurement error in the former estimator. Card (2001) alternatively proposes the potential heterogeneity of returns to schooling across individuals in a population as an explanation for why the IV estimator is greater compared to the OLS one. However, Card's explanation applies to instruments based on a policy-change or supply-side intervention while this study uses family background and distance to school as instruments suggesting that his explanation may not be relevant at least in this case.

[15] From the original sample from which the household fixed-effects sample was derived we deleted observations on non-biological children, fathers/mothers-in-law, brothers/sisters-in-law, other relatives and servants,

creates a potential sample selection problem, the procedure allows us to net out unobserved heterogeneity in a family differenced model. Although unobserved traits are not identical across family members, it is likely they are much more similar within a family than across families and, as such, family fixed effects estimation gives an estimate of the return to education that reduces endogeneity bias without necessarily eliminating it entirely[16].

Table 13 presents family fixed-effect findings. As before, the rate of return to an additional year of schooling is roughly 5 per cent (compared to about 5.6 per cent using IV in Table 12A and compared to about 5.5 per cent using OLS in Table 9). The similarity of the estimated RORE across OLS, IV and family fixed effects techniques suggests that ability bias is not strong. Secondly, there are direct returns to schooling rather than to human capital which hints at credentialism and (based on previous evidence) at rewards to non-cognitive traits such as attitude and behaviour.

[Table 13 about here]

5. Conclusions

This study investigates the relationship between dimensions of human capital – years of schooling, cognitive skills and ability – on the one hand, and occupational attainment and earnings, on the other. The aim was to ask whether one or more of these dimensions of human capital raise earnings within any given occupation and/or raise earnings indirectly via facilitating entry into well paying occupations. The key objective, therefore, was to ask whether evidence from Pakistan supports the human capital, credentialist or signalling hypotheses.

There are several interesting findings. Firstly, the evidence shows that while basic-order cognitive skills (literacy) promote women's entry into the lucrative wage occupations, for men higher order cognitive skills are required for the same benefit. Much of the direct effect of cognitive skills disappears after conditioning on schooling, i.e. skills do not have an independent effect on earnings over and above the effect via schooling. Secondly, there is no direct effect of measured ability in the earnings functions, either for men or women. That OLS, IV and household fixed-effect estimates consistently reveal an economic rate of return of about 5 per cent suggests (in line with much of the international literature) that 'ability bias' is not strong, and there appears to be no evidence of signalling. Thirdly, there is a direct return to schooling for men while for women there is a return to cognitive skills, though in IV estimates the effect is not statistically significant possibly due to small sample size. The data clearly support the human capital hypothesis for women even on a small sample. While they suggest a credentialist hypothesis for men – since skills do not have an

---

employees or other non-relatives of the household head reported to be resident in the household. This resulted in a loss of 19 individual observations.

[16] We are unable to report fixed-effects estimates separately for males and females due to sample size constraints. We also estimated fixed-effects with schooling represented as quadratic but the linear model is preferred as schooling and schooling2 are not significant in the quadratic specification. Models were also estimated with interaction terms (Schooling*male) and (Literate*Male) to determine gender effects but these too were insignificant.

independent effect on earnings for men, over and above the effect via schooling – we cannot be confident of this conclusion due to the bigger sample selection problem for men (a much higher proportion of waged men than women had missing cognitive skills scores). Future studies with better data on cognitive skills would need to confirm it for us to be confident of the rather pessimistic conclusion of credentialism for Pakistani males in wage employment.

From a policy perspective, these findings suggest that for waged women, education raises earnings and (if workers are paid their marginal product) raises worker productivity, suggesting an efficiency rationale for public funding of education. For men there is a less clear story, as there is a hint of credentialism.

Education among both men and women is highly correlated with positive behaviour traits and it seems that education also affects earnings/productivity via its effects on these positive traits which are rewarded in the labour market.

The paper also examined the shape of the education-earnings relationship in Pakistan. While it has conventionally been assumed/found that earnings increase with education at a decreasing rate, i.e. the relationship is concave (Psacharopoulos, 1994), data for Pakistan for the late 1990s suggested that the relationship is convex (Aslam, 2008; Aslam, Kingdon and Soderbom, 2008), i.e. that the return to an extra year of education progressively increases with education level. The results from the present study also confirm the pronounced convexity of the education-earnings relationship.

References

Alderman, H., Behrman, J.R., Ross, D.R. & Sabot, R. (1996). The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market. *Oxford Bulletin of Economics and Statistics*, 58 (1), pp. 29-55.

Aslam, M. (2008). Education Gender Gaps in Pakistan: Is the Labour Market to Blame?, forthcoming in *Economic Development and Cultural Change.*

Aslam, M., Kingdon, G.G. & Mans Söderbom, (2008). Is Education a Path to Gender Equality in the Labor Market? Evidence from Pakistan. In *Education as a Path to Gender Equality*, forthcoming (The World Bank).

Behrman, J.R., Ross, D. & Sabot, R. (???). Improving the Quality Versus Increasing the Quantity of Schooling: Evidence for Rural Pakistan. PIER Working Paper 02-022.

Bishop, J.H. (1991). The Impact of Academic Competencies on Wages, Unemployment and Job Performance. Center for Advanced Human Resource Studies (CAHRS) Working Paper No. 91-34, Cornell University.

Boissiere, M., Knight, J. B. & Sabot, R.H. (1985). Earnings, Schooling, Ability, and Cognitive Skills. *The American Economic Review*, 75 (5), pp. 1016-1030.

Card. D. (2001). Estimating the Returns to Schooling: Progress in some Persistent Econometric Problems. *Econometrica*, 69 (5), pp. 1127-60.

Cawley, J., Heckman, J. & Vytlacil, E. (2001). Three Observations on Wages and Measured Cognitive Ability. *Labour Economics*, 8, pp. 419-442.

Currie, J. & Thomas, D. (2001). Early Test Scores, School Quality and SES: Longrun Effects on Wage and Employment Outcomes. *Worker Wellbeing in a Changing Labor Market*, 20, pp. 103-132.

Glewwe, P. (1996). The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment. *Journal of Development Economics*, 51, pp. 267-90.

Griliches, Z. (1977). Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*, 45 (1), pp. 1-22.

Hanushek, E.A. (2005). The Economics of School Quality. *German Economic Review*, 6 (3), pp. 269-286.

Hanushek, E.A. & Zhang, L. (2006). Quality Consistent Estimates of International Returns to Skill. National Bureau of Economic Research, WP12664, Cambridge, MA, NBER, November.

Hanushek, E.A. & Woessmann, L. (2007). The Role of School Improvement in Economic Development. Working Paper PEPG 07-01, Program of Education Policy and Governance, Kennedy School of Government, Harvard University.

Ishikawa, M. & Ryan, D. (2002). Schooling, Basic Skills and Economic Outcomes. *Economics of Education Review*, 21, pp. 231-243.

Jolliffe, D. (1998). Skills, Schooling and Household Income in Ghana. *The World Bank Economic Review*, 12 (1), pp. 81-104.

Katz, A. (1993). Comments on Policy Implications. In *Conference on Literacy and the American Worker: Implications for Public Policy, College of William and Mary, Williamsburg, VA.*

Kingdon, G.G. & Söderbom, M. (2007). Education, Skills and Labor Market Outcomes: Evidence from Pakistan. University of Oxford (Mimeo).

Murnane, R.J., Willett, J.B. & Levy,F. (1995). The Growing Importance of Cognitive Skills in Wage Determination. NBER Working Paper No. 5076 (March).

Psacharopoulos (1994). Returns to Investment in Education: A Global Update. *World Development*, 22 (9), pp. 1325-1343.

Riboud, M., Savchenko, Y. & Tan, H. (2006). The Knowledge Economy and Education and Training in South Asia: A Mapping Exercise of Available Survey Data. World Bank Working Paper, South Asia Region, 2006.

Rivera-Batiz, F. L (1992). Quantitative Literacy and the Likelihood of Employment among Young Adults in the United States. *The Journal of Human Resources*, 27 (2), pp. 313-328.

Söderbom, M., Teal, F., Wambugu, A. & Kahyarara, G. (2005). The Dynamics of Returns to Education in Kenyan and Tanzanian Manufacturing. Global Poverty Research Group Working Paper Series, No. 017.

Trostel, P., Walker, I. & Woolley, P. (2002). Estimates of the Economic Return to Schooling for 28 Countries. *Labour Economics*¸ 9, pp. 1-16.

Tyler, J.H. (2002). Basic Skills and the Earnings of Dropouts. Brown University Department of Economics Working Paper No. 2002-09 (November).

Appendix Tables

A1: Effect of Ravens on Log Earnings in Earnings Functions – various specifications

Log earnings	(1)	(2)	(3)	(4)	(5)	(6)
	ALL		**Male**		**Female**	
Male	0.577	0.618				
	(6.07)**	(6.30)**				
Schooling	0.048	0.013	0.043	0.009	0.076	0.027
	(7.81)**	(0.82)	(7.27)**	(0.60)	(4.60)**	(0.49)
Schooling2		0.003		0.003		0.003
		(2.30)*		(2.46)**		(0.94)
Exp	0.078	0.077	0.061	0.059	0.206	0.205
	(10.45)**	(10.30)**	(9.28)**	(9.04)**	(6.21)**	(6.10)**
Exp2	-0.002	-0.002	-0.001	-0.001	-0.006	-0.006
	(8.48)**	(8.47)**	(7.39)**	(7.29)**	(4.87)**	(4.75)**
Ravens	0.011	0.010	0.007	0.004	0.035	0.041
	(1.02)	(0.93)	(0.69)	(0.44)	(1.11)	(1.37)
Constant	6.907	6.940	7.654	7.739	5.949	5.945
	(49.16)**	(49.54)**	(84.77)**	(84.85)**	(16.71)**	(16.65)**
Observations	686	686	559	559	127	127
R-squared	0.29	0.29	0.22	0.23	0.34	0.35

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% or better.

A2: Effect of Basic-Order Literacy Dummies on Log Earnings, by Gender

Log earnings	(1)	(2)	(3)	(4)
	Male		**Female**	
Schooling	0.057	0.024	0.072	-0.006
	(9.92)**	(1.72)*	(1.76)*	(0.07)
Schooling2		0.002		0.005
		(2.51)*		(1.08)
Exp	0.052	0.050	0.199	0.199
	(8.34)**	(8.07)**	(5.89)***	(5.83)***
Exp2	-0.001	-0.001	-0.005	-0.005
	(6.57)**	(6.54)**	(4.60)**	(4.54)**
Dumlit1	-0.292	-0.266	0.000	0.000
	(1.87)*	(1.74)*	(.)	(.)
Dumlit2	-0.131	-0.096	0.487	0.724
	(1.06)	(0.78)	(0.98)	(1.18)
Dumlit3	-0.283	-0.239	0.019	0.278
	(2.34)**	(1.90)*	(0.03)	(0.42)
Dumlit4	-0.275	-0.238	0.083	0.261
	(3.41)***	(2.92)***	(0.16)	(0.42)
Dumlit5	-0.203	-0.194	0.222	0.419
	(3.42)***	(3.27)***	(0.43)	(0.69)
Constant	7.850	7.910	6.224	6.261
	(141.62)***	(134.18)***	(22.46)***	(22.46)***
Observations	836	836	132	132
R-squared	0.18	0.18	0.32	0.32

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% or better

A3: Effect of Higher-Order Skills on Log Earnings, by Gender

Log earnings	**(1)** **Male**	**(2)**	**(3)** **Female**	**(4)**
Schooling	0.060	0.021	0.074	0.003
	(6.41)***	(1.14)	(2.20)**	(0.04)
Schooling2		0.003		0.004
		(2.28)**		(0.98)
Exp	0.062	0.060	0.211	0.209
	(9.33)***	(8.95)**8	(6.34)***	(6.21)***
Exp2	-0.001	-0.001	-0.006	-0.006
	(7.47)**8	(7.28)***	(4.89)***	(4.76)***
Tlit	-0.004	0.023	0.172	0.248
	(0.09)	(0.55)	(1.10)	(1.45)
Tlit2	-0.001	-0.005	-0.015	-0.025
	(0.22)	(0.83)	(0.84)	(1.27)
Tmaths	-0.035	-0.017	-0.119	-0.090
	(1.39)	(0.64)	(0.98)	(0.71)
Tmaths2	0.002	0.000	0.010	0.007
	(0.90)	(0.18)	(0.98)	(0.73)
Constant	7.714	7.765	6.155	6.199
	(123.60)***	(117.38)***	(22.33)***	(22.05)***
Observations	559	559	127	127
R-squared	0.22	0.23	0.35	0.35

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better

Table A4 – Coefficients on Schooling and Skills from Wage Functions, by Worker Group

Log earnings	**Male**				**Female**			
	Schooling		**Literate**		**Schooling**		**Literate**	
Occupation								
Elementary	0.035	***	0.034		0.014		0.653	
	(2.33)		(0.29)		(0.35)		(1.03)	
Blue-collar	0.048	***	-0.170		-0.123		2.099	
	(3.41)		(1.10)		(0.55)		(1.32)	
White-collar	0.126	***	---		0.165	***	---	
	(4.57)				(3.68)			
Region								
Rural	0.056	***	-0.121		-0.009		1.182	*
	(4.99)		(1.16)		(1.15)		(1.79)	
Urban	0.047	***	-0.168		0.195	***	-1.478	**
	(3.41)		(1.14)		(4.42)		(2.01)	
Cohort								
Young (<=30)	0.012		-0.094		-0.047		1.564	*
	(0.70)		(0.65)		(0.73)		(1.89)	
Old (>30)	0.064	***	-0.105		0.146	***	-0.649	
	(6.48)		(1.01)		(3.73)		(1.29)	
Empl. Sector								
Private	0.025	**	0.018		0.006		0.712	
	(1.97)		(0.18)		(0.09)		(0.83)	
Govt.	0.075	***	-0.366	*	0.114	***	-0.837	*
	(6.16)		(1.84)		(3.19)		(-1.92)	
Payment Type								
Regular	0.073	***	-0.254	**	0.132	***	-0.016	
	(7.37)		(2.05)		(3.54)		(0.03)	
Casual	0.001		0.215	*	0.035		1.123	
	(0.09)		(1.92)		(0.45)		(1.12)	

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better.

Table A5: Coefficients on Schooling, Schooling2 and Skills from Wage Functions, by Worker Group

Log earnings	Male Schooling	Schooling2	Literate	Female Schooling	Schooling2	Literate
Occupation						
Elementary	-0.014	0.004	0.145	0.048	-0.002	0.568
	(0.42)	(1.42)	(1.27)	(0.61)	(0.49)	(0.79)
Blue-collar	0.052	-0.000	-0.184	-0.276	0.016	2.285
	(1.57)	(0.11)	(1.08)	(0.95)	(0.58)	(1.58)
White-collar	0.073	0.002	---	-0.276	0.017	---
	(0.42)	(0.30)		(0.59)	(0.96)	
Region						
Rural	0.008	0.003	0.036	-0.329 **	0.016 **	2.604 ***
	(0.31)	(1.64)	(0.34)	(2.38)	(2.37)	(3.33)
Urban	-0.038	0.005 **	0.108	0.272	-0.003	-1.900
	(0.99)	(2.42)	(0.58)	(1.09)	(0.30)	(1.37)
Cohort (<=30, >30)						
Young	-0.025	0.002	0.010	-0.805 ***	0.038 **	4.385 ***
	(0.63)	(0.84)	(0.07)	(3.08)	(2.96)	(4.08)
Old	0.012	0.003 **	0.081	0.050	0.004	-0.149
	(0.48)	(2.02)	(0.68)	(0.36)	(0.70)	(0.19)
Employ. Sector						
Private	0.014	0.001	0.045	-0.405 **	0.021 **	2.447 **
	(0.47)	(0.29)	(0.45)	(2.17)	(2.45)	(2.28)
Govt.	-0.002	0.003	0.040	0.051	0.003	-0.539
	(0.03)	(1.12)	(0.11)	(0.18)	(0.23)	(0.41)
Payment Type						
Regular	0.028	0.002	-0.073	0.070	0.003	0.311
	(1.00)	(1.54)	(0.56)	(0.54)	(0.48)	(0.46)
Casual	0.012	-0.001	0.198 *	-0.353	0.019	2.632 **
	(0.38)	(0.35)	(1.85)	(1.34)	(1.64)	(2.01)

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better.

Tables

Table1: Distribution of the Labour Force by Gender (15-60)

Labour Force Status	Male		Female		All	
	N	%	N	%	N	%
Out of the Labour Force	407	17.7	1559	69.6	1966	43.2
Unemployed	120	5.2	201	8.9	321	7.1
Agriculture Self Employed	370	16.1	208	9.2	578	12.7
Non-Agriculture Self Employed	439	19.1	121	5.4	560	12.3
Wage Employed	961	42.0	169	7.5	1130	25.0
All	2297	100	2258	100	4555	100

Table2: Completed education levels for males and females (15-60)

Education Level	Description	Males		Females	
		N	%	N	%
NO_EDU	No education	220	23.0	68	40.2
LESSPRIM	Less than primary – includes 'kacchi', 'pakki', 1,2,3 and 4 completed years of schooling	58	6.0	4	2.4
PRIMARY	Completed 5, 6 or 7 years of schooling	157	16.4	8	4.7
MIDDLE	Completed 8 or 9 years of schooling	159	16.6	7	4.1
MATRIC	Completed 10 or 11 years of schooling	221	23.0	29	17.2
INTER	Completed 12 or 13 years of schooling	78	8.1	19	11.2
BA_MORE	Completed 14 or more years of schooling	67	7.0	34	20.1
ALL		960	100	169	100

Table 3: Key summary statistics (Wage earners, aged 15-60)

Variable	Mean Years of Education (a)	t-value of difference in (a)	% Literate (b)	t-value of difference in (b)	% Numerate (c)	t-value of difference in (c)	Mean tscore (d)	t-value of difference in (d)	Mean conditional earnings (Rs./month) (e)	t-value of difference in (e)	Median conditional earnings (Rs./month) (f)
Gender											
Male	6.7	0.5	0.69	-2.2	0.88	1.9	7.7	-1.7	6066	-3.2	4500
Female	6.9		0.59		0.94		6.7		4417		2400
Region											
Urban	8.4	7.6	0.85	7.5	0.95	3.5	9.9	7.3	5970	0.5	5000
Rural	6.1		0.58		0.87		6.4		5785		4500
Sector of Employment											
Private	5.5	-14.8	0.58	-9.5	0.87	-4.0	5.9	-11.6	5224	-6.0	4000
Government	10.3		0.95		0.98		12.1		7635		6000
School Type											
Private	11.1	3.2	1.00	1.4	1.00	0.6	12.9	2.0	9803	3.1	4500
Government	9.1		0.94		0.98		10.6		6190		5000
Age Cohort											
Old (>30)	6.7	-0.9	0.63	-3.4	0.89	-0.6	7.5	0.1	6280	3.4	5000
Young (<=30)	6.9		0.75		0.90		7.5		5026		3900
Payment type											
Daily	4.5	-13.7	0.48	-9.9	0.84	-4.5	4.6	-11.0	4997	-3.3	4500
Regular	8.3		0.81		0.94		9.7		6252		5000
Occupation											
Elementary	4.3		0.46		0.84		4.5		5019		4000
Blue-collar	7.5		0.79		0.92		8.6		6278		5000
White-collar	12.4		1.00		1.00		13.5		6989		6000

Note: Literate = 1 if individual scored 1 or more (out of 5) in the short literacy test, Numerate = 1 if individual scored 3 or more (out of 5) in the short maths test. Shaded cells represent significance at 10% or better.

Table 4A: Marginal Effects of Multinomial Logit of Activity Status – Conditioning on Basic Order Skills

	Male					**Female**				
	OLF	Unempl	Agri Self Employed	Non-agri Self Employed	Wage Employed	OLF	UnEmpl.	Agri Self Employed	Non-agri Self Employed	Wage Employed
Age	-0.0059	-0.0004	0.0036	0.0016	0.0012	0.0000	-0.0006	-0.0005	0.0001	0.0009
	(-5.07)***	(-0.67)	(3.47)***	(1.39)	(0.94)	(0.06)	(-1.04)	(-1.06)	(0.31)	(1.83)*
Head	-0.0221	-0.0312	-0.0620	0.0834	0.0319	-0.0506	0.0305	-0.0452	0.0540	0.0112
	(-0.99)	(-2.30)**	(-2.43)**	(2.56)**	(0.93)	(-0.70)	(0.58)	(-2.03)**	(1.11)	(0.28)
Literate	0.1295	0.0228	-0.1165	0.0283	-0.0641	0.0372	0.0077	-0.1042	0.0141	0.0451
	(8.04)***	(1.94)**	(-4.74)***	(1.17)	(-2.31)**	(1.63)	(0.53)	(-8.07)***	(1.17)	(3.34)***
Ravens	-0.0008	-0.0021	-0.0067	0.0099	-0.0003	-0.0092	0.0055	-0.0029	0.0010	0.0057
	(-0.27)	(-1.10)	(-1.84)*	(2.60)***	(-0.07)	(-2.30)**	(2.05)**	(-1.47)	(0.51)	(2.63)***
N	1749					2106				

Note: Robust z-statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better; Base category = Out of the Labour Force. Literate = 1 if individual scored greater than 0 (out of 5) in the short literacy test. Ravens = total score on ravens test (max score = 20).

Table 4B: Marginal Effects of Multinomial Logit of Activity Status – Conditioning on Basic Order Skills and Schooling

	Male					**Female**				
	OLF	Unempl.	Agri Self Employed	Non-agri Self Employed	Wage Employed	OLF	Unempl.	Agri Self Employed	Non-agri Self Employed	Wage Employed
Age	-0.0058	-0.0004	0.0039	0.0023	0.0000	0.0004	-0.0007	-0.0006	0.0000	0.0009
	(-4.83)***	(-0.68)	(3.77)***	(1.98)*	(0.02)	(0.39)	(-1.20)	(-1.50)	(0.09)	(1.82)
Head	-0.0221	-0.0306	-0.0678	0.0754	0.0452	-0.0589	0.0343	-0.0390	0.0578	0.0059
	(-0.97)	(-2.21)**	(-2.64)***	(2.28)**	(1.31)	(-0.81)	(0.63)	(-1.92)*	(1.14)	(0.16)
Schooling	0.0086	0.0009	0.0080	0.0265	-0.0438	0.0311	-0.0122	-0.0052	-0.0065	-0.0072
	(0.95)	(0.17)	(-2.33)**	(2.71)***	(-4.13)***	(3.11)***	(-1.96)*	(-0.94)	(-1.13)	(-1.62)
Schooling2	-0.0004	0.0001	-0.0013	-0.0020	0.0036	-0.0022	0.0010	-0.0006	0.0004	0.0013
	(-0.71)	(-0.18)	(-2.36)**	(-3.01)***	(5.45)***	(-3.01)***	(2.48)**	(-1.24)	(1.02)	(4.53)***
Literate	0.1072	0.0145	-0.0833	-0.0152	-0.0231	-0.0081	0.0257	-0.0298	0.0342	-0.0220
	(3.90)***	(0.78)	(-0.92)	(-0.40)	(-0.51)	(-0.21)	(0.95)	(-1.74)*	(1.47)	(-1.37)
Ravens	-0.0009	-0.0025	-0.0035	0.0131	-0.0061	-0.0062	0.0051	-0.0016	0.0012	0.0016
	(-0.31)	(-1.25)		(3.10)***	(-1.38)	(1.58)	(1.90)*	(-0.89)	(0.58)	(0.81)
N	1749					2106				

Note: Robust z-statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better; Base category = Out of the Labour Force. Literate = 1 if individual scored greater than 0 (out of 5) in the short literacy test. Ravens = total score on ravens test (max score = 20).

Table 5A: Marginal Effects of Multinomial Logit of Activity Status – Conditioning on Higher Order Skills

	Male					Female				
	OLF	UnEmpl.	Agri Self Employed	Non-agri Self Employed	Wage Employed	OLF	UnEmpl.	Agri Self Employed	Non-agri Self Employed	Wage Employed
Tlit	0.0228	0.0128	-0.0026	0.0063	-0.0392	0.0184	-0.0135	-0.0231	0.0143	0.0040
	(1.62)	(1.57)	(-0.16)	(0.35)	(-1.95)*	(1.00)	(-1.16)	(-1.83)*	(1.60)	(0.45)
Tlit2	-0.0021	-0.0006	-0.0009	-0.0018	0.0053	0.0011	0.0012	0.0000	-0.0023	-0.0001
	(-1.19)	(-0.56)	(-0.37)	(-0.74)	(1.96)**	(0.39)	(0.71)	(0.01)	(-1.75)*	(-0.04)
Tmaths	0.0172	0.0045	-0.0121	0.0173	-0.0269	-0.0176	0.0168	-0.0077	0.0012	0.0072
	(1.82)*	(0.88)	(1.16)	(1.52)	(-2.06)**	(-1.32)	(2.07)**	(-0.88)	(0.19)	(1.20)
Tmaths2	-0.0010	-0.0009	0.0003	-0.0012	0.0028	0.0005	-0.0015	0.0006	0.0002	0.0002
	(-1.30)	(-1.79)*	(0.30)	(-1.16)	(2.32)**	(0.36)	(-1.68)*	(0.57)	(0.29)	(0.30)
Ravens	-0.0008	-0.0013	-0.0055	0.0103	-0.0028	-0.0059	0.0065	-0.0027	0.0001	0.0020
	(-0.27)	(-0.63)	(-1.45)	(2.54)**	(-0.63)	(-1.45)	(2.32)**	(-.138)	(0.05)	(0.97)
N	1746					2101				

Note: Robust z-statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better; Base category = Out of the Labour Force. Controls also include HEAD (1 if individual is household head, 0 otherwise) and age (years). Tlit = total score on literacy test (max = 8), Tlti2 = Tlit squared, Tmaths = total score on maths test (max = 12), Tmaths2 = Tmaths squared. Ravens = total score on ravens test (max score = 20).

Table 5B: Marginal Effects of Multinomial Logit of Activity Status – Conditioning on Higher Order Skills and Schooling

	Male					Female				
	OLF	UnEmpl.	Agri Self Employed	Non-agri Self Employed	Wage Employed	OLF	UnEmpl.	Agri Self Employed	Non-agri Self Employed	Wage Employed
Schooling	0.0190	0.0007	0.0024	0.0182	-0.0403	0.0338	-0.0056	-0.0085	-0.0073	-0.0125
	(2.23)**	(0.16)	(0.30)	(1.86)*	(-3.96)***	(3.73)***	(-1.04)	(1.60)	(-1.37)	(-2.90)***
Schooling2	-0.0009	0.0001	-0.0010	-0.0016	0.0034	-0.0024	0.0009	-0.0004	0.0004	0.0015
	(-1.78)*	(0.44)	(-1.67)*	(-2.45)**	(5.16)***	(-3.23)***	(2.40)**	(-0.81)	(0.86)	(5.13)***
Tlit	0.0118	0.0107	0.0015	-0.0001	-0.0239	-0.0123	-0.0168	0.0014	0.0233	0.0044
	(0.83)	(1.22)	(0.08)	(0.00)	(-1.10)	(-0.59)	(-1.22)	(0.11)	(2.25)**	(0.43)
Tlit2	-0.0009	-0.0005	-0.0008	-0.0007	0.0029	0.0050	0.0011	-0.0017	-0.0033	-0.0011
	(-0.53)	(-0.42)	(-0.33)	(-0.29)	(1.02)	(1.58)	(0.57)	(-0.71)	(-2.21)**	(-0.85)
Tmaths	0.0089	0.0038	-0.0120	0.0087	-0.0094	-0.0240	0.0151	0.0010	0.0042	0.0036
	(0.92)	(0.71)	(-1.10)	(0.69)	(-0.66)	(-1.77)*	(1.70)*	(0.13)	(0.59)	(0.54)
Tmaths2	-0.0005	-0.0009	0.0007	-0.0003	0.0010	0.0015	-0.0017	0.0003	0.0000	0.0000
	(-0.67)	(-1.81)*	(0.69)	(-0.27)	(0.79)	(1.00)	(-1.81)*	(0.26)	(-0.05)	(0.00)
Ravens	-0.0008	-0.0017	-0.0038	0.0126	-0.0063	-0.0056	0.0061	-0.0019	0.0003	0.0010
	(-0.26)	(0.82)	(-0.98)	(2.94)***	(-1.42)	(-1.37)	(2.21)**	(-1.06)	(0.15)	(0.49)
N	1746					2101				

Note: Robust z-statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better; Base category = Out of the Labour Force. Controls also include HEAD (1 if individual is household head, 0 otherwise) and age (years). Tlit = total score on literacy test (max = 8), Tlti2 = Tlit squared, Tmaths = total score on maths test (max = 12), Tmaths2 = Tmaths squared. Ravens = total score on ravens test (max score = 20).

Table 6: Marginal Effects from Logit of Sectoral Attainment (Higher Order Skills) - dependent variable (=1 if in government sector job, 0 if in private) . Sample of wage employees only.

Govt Job	(1) **Males**	(2) **Females**
Schooling	0.0468	0.0973
	(3.47)***	(2.76)***
Schooling2	-0.0004	-0.0037
	(-0.49)	(-2.57)**
Tlit	0.0007	0.0079
	(0.03)	(0.19)
Tlit2	-0.0020	0.0027
	(-0.57)	(0.50)
Tmaths	-0.0464	-0.0584
	(-2.44)**	(1.41)
Tmaths2	0.0050	0.0043
	(3.17)***	(1.36)
N	560	131

Note: Robust z statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better; Controls also include Age (yrs) and Head (=1 if individual is household head, 0 otherwise). Raven score insignificant and hence dropped. Tlit = total score on literacy test (max = 8), Tlti2 = Tlit squared, Tmaths = total score on maths test (max = 12), Tmaths2 = Tmaths squared. Literacy and numeracy test scores were available only for a sub-sample of waged workers, that is why N here is lower than that in Table 1.

Table 7: Marginal Effects from Multinomial Logit of Occupational Attainment (Higher Order Skills) (= 0 if Elementary, = 1 if Blue-Collar and = 2 if White-Collar). Sample of wage employees only.

	Male			**Female**		
	Elementary	Blue- Collar	White- Collar	Elementary	Blue- Collar	White- Collar
Schooling	-0.0336	0.0259	0.0077	-0.3262	0.0722	0.2540
	(-1.58)	(1.22)	(1.05)	(-2.04)**	(1.66)*	(1.43)
Schooling2	0.0004	-0.0002	-0.0002	0.0170	-0.0067	-0.0103
	(0.28)	(-0.15)	(-1.44)	(2.83)***	(-2.04)**	(-1.40)
Tlit	-0.0160	0.0162	-0.0002	0.0814	-0.0178	-0.0636
	(-0.39)	(0.39)	(-0.09)	(0.54)	(-0.20)	(-0.71)
Tlit2	0.0022	-0.0024	0.0001	-0.0077	-0.0003	0.0080
	(0.41)	(-0.43)	(0.33)	(-0.37)	(-0.02)	(0.78)
Tmaths	0.0152	-0.0120	-0.0032	-0.0086	-0.1124	0.1209
	(0.54)	(-0.43)	(-0.57)	(-0.07)	(-1.91)*	(1.25)
Tmaths2	-0.0035	0.0031	0.0003	-0.0041	0.0138	-0.0097
	(-1.31)	(1.19)	(0.61)	(-0.32)	(1.78)*	(-1.23)
N	542			131		

Notes: As in Table 6.

Table 8: Earnings Functions – Various Specifications

Log earnings	(1) **All**	(2)	(3) **Male**	(4)	(5) **Female**	(6)
Male	0.581	0.622				
	(6.11)***	(6.35)***				
Schooling	0.051	0.015	0.045	0.009	0.083	0.049
	(9.06)***	(0.94)	(8.52)***	(0.64)	(5.27)***	(0.87)
Schooling2		0.003		0.003		0.002
		(2.35)**		(2.49)**		(0.65)
Exp	0.078	0.077	0.061	0.059	0.205	0.205
	(10.45)***	(10.30)***	(9.27)***	(9.01)***	(6.21)***	(6.12)***
Exp2	-0.002	-0.002	-0.001	-0.001	-0.006	-0.006
	(8.47)***	(8.46)***	(7.37)**8	(7.27)***	(4.87)***	(4.77)***
Constant	6.975	7.003	7.702	7.770	6.188	6.217
	(55.88)***	(56.37)***	(123.53)***	(120.01)***	(22.37)***	(22.11)***
Observations	686	686	559	559	127	127
R-squared	0.29	0.29	0.22	0.23	0.34	0.34

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better

Table 9: Earnings Functions, by gender (with Basic Order Skills)

Log earnings	(1) **All**	(2)	(3) **Male**	(4)	(5) **Female**	(6)
Male	0.590	0.613	-		-	
	(6.09)***	(6.27)***				
Schooling	0.055	-0.025	0.052	-0.008	0.057	-0.295
	(6.01)***	(-0.98)	(6.05)***	(0.38)	(1.34)	(2.07)**
Schooling2		0.004		0.004		0.016
		(3.09)***		(2.71)***		(2.75)***
Exp	0.078	0.077	0.061	0.059	0.206	0.201
	(10.43)***	(10.23)***	(9.25)***	(8.94)***	(6.26)***	(6.08)***
Exp2	-0.002	-0.002	-0.001	-0.001	-0.006	-0.005
	(-8.50)***	(-8.23)***	(7.41)***	(7.15)***	(4.91)***	(4.66)***
Literate	-0.054	0.226	-0.103	0.098	0.347	2.047
	(-0.56)	(1.96)**	(1.22)	(1.04)	(0.64)	(2.38)**
Constant	6.981	7.000	7.729	7.763	6.162	6.229
	(55.75)***	(56.51)***	(121.71)***	(119.51)***	(22.32)***	(22.62)***
N	686	686	559	559	127	127
R-squared	0.29	0.29	0.22	0.23	0.34	0.37

Notes: Robust t statistics in parentheses

Table 10: Earnings Functions, by Gender and Region

Log earnings	(1) **Male** **Rural**	(2) **Urban**	(3) **Female** **Rural**	(4) **Urban**	(5) **Male** **Rural**	(6) **Urban**	(7) **Female** **Rural**	(8) **Urban**
Schooling	0.056	0.047	-0.009	0.195	0.008	-0.038	-0.329	0.272
	(4.99)***	(3.41)***	(0.15)	(4.42)***	(0.31)	(0.99)	(2.38)**	(1.09)
Schooling2					0.003	0.005	0.016	-0.003
					(1.64)	(2.42)**	(2.37)**	(0.30)
Literate	-0.121	-0.168	1.182	-1.478	0.036	0.108	2.604	-1.900
	(1.16)	(1.14)	(1.79)*	(2.01)**	(0.34)	(0.58)	(3.33)***	(1.37)
Observations	391	168	78	49	391	168	78	49
R-squared	0.20	0.27	0.37	0.45	0.20	0.29	0.40	0.45

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better; Controls for Experience and its square included but not shown.

Table 11 – Effects of Socialisation Skills and Personality traits (when aged 15) on log earnings in adulthood

Log earnings	(1) Regression coefficient not conditioning on SCHOOLING and LITERATE	(2) Regression coefficient conditioning on SCHOOLING	(3) Regression coefficient conditioning on LITERATE	(4) Regression coefficient conditioning on SCHOOLING and LITERATE
Male	0.569	0.592	0.495	0.581
	(5.62)***	(6.18)***	(5.14)***	(5.90)***
Exp	0.075	0.079	0.079	0.080
	(9.57)***	(10.47)***	(10.12)***	(10.50)***
Exp2	-0.002	-0.002	-0.002	-0.002
	(8.22)***	(8.52)***	(8.19)***	(8.55)***
Postrait	0.061	0.013	0.016	0.002
	(2.39)**	(0.54)	(0.65)	(0.09)
Negtrait	0.006	-0.003	-0.003	-0.007
	(0.15)	(0.08)	(0.08)	(0.19)
Religious	0.101	0.033	0.035	0.013
	(1.77)*	(0.59)	(0.61)	(0.24)
Schooling	-	0.048	-	0.054
		(7.96)***		(5.72)***
Literate	-	-	0.436	-0.048
			(6.80)***	(0.50)
Constant	7.183	6.948	7.023	6.973
	(53.38)***	(51.37)***	(50.68)***	(50.52)***
Observations	700	700	675	675
R-squared	0.21	0.29	0.26	0.29

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better; **Postrait** (positive personality traits) = made friends easily (equals 1 if did, 0 otherwise) + had a lot to say (equals 1 if did, 0 otherwise) + made plans and stuck to them (equals 1 if did, 0 otherwise) + accepted people as they were (equals 1 if did, 0 otherwise); **Negtrait** (negative personality traits) = found it difficult to get down to work equals 1 if did, 0 otherwise) + did just enough work to get by (equals 1 if did, 0 otherwise); **Religious** = 1 if was sympathetic to religious parties/causes, 0 otherwise.

Table 12A: OLS and Instrumental Variable Estimates of Log Earnings (All, i.e. Male and Female)

Log earnings	**<u>All</u>** **OLS** **(1)**		**IV** **(2)**		**First Stage** **(3)**			
					Schooling		**Literate**	
Schooling	0.056	***	0.023		-		-	
	(6.03)		(0.32)					
Literate	-0.047		0.503		-		-	
	(-0.49)		(0.53)					
Male	0.590	***	0.504	***	-0.217		0.132	***
	(5.99)		(2.79)		(-0.52)		(3.57)	
Exp	0.078	***	0.079	***	0.004		-0.001	
	(10.21)		(9.90)		(0.09)		(-0.32)	
Exp2	-0.002	***	-0.002	***	-0.002		-0.000	*
	(-8.36)		(-7.55)		(-1.47)		(-1.89)	
Pdist*1000	-		-		-0.511	***	-0.037	***
					(-4.01)		(-2.90)	
Pdist2*1000	-		-		0.000	*	0.000	
					(1.87)		(1.37)	
Fedu	-		-		0.460	***	0.036	***
					(11.07)		(10.63)	
Medu	-		-		0.214	***	0.003	
					(2.85)		(0.60)	
Ravens	-		-		0.580	***	0.043	***
					(10.41)		(8.68)	
Constant	6.980	***	6.851		1.375	*	0.218	***
	(54.84)		(41.12)		(1.93)		(3.35)	
R2	0.287		0.255		0.439		0.342	
N	667		667		667		667	
P-value (F test of excluded instruments)	-		-		0.000		0.000	
P-value (overid test)	-		-		0.900			

Note: Robust t statistics are in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better; **Pdist** = distance in meters to nearest primary school when individual was of school-going age, **Pdist2** = Pdist squared, **Fedu** = Father's completed education in years, **Medu** = Mother's completed education in years, **Ravens** = score on the Ravens test (max = 20).

Table 12B: OLS and Instrumental Variable Estimates of Log Earnings (Males and Females)

Log earnings	Males OLS (1)	IV (2)	First Stage (3) Schooling	Literate	Females OLS (1)	IV (2)	First Stage (3) Schooling	Literate
Schooling	0.053 ***	0.102 *	-	-	0.058	0.067	-	-
	(6.07)	(1.69)			(1.34)	(0.60)		
Literate	-0.097	-0.796	-	-	0.345	0.504	-	-
	(-1.13)	(1.00)			(0.63)	(0.34)		
Exp	0.060 ***	0.058 ***	0.034	-0.001	0.206 ***	0.216 ***	-0.264 **	-0.020 *
	(9.08)	(7.72)	(0.78)	(-0.17)	(6.19)	(6.97)	(-2.17)	(-1.92)
Exp2	-0.001 ***	-0.001 ***	-0.003 **	-0.000 **	-0.006 ***	-0.006 ***	0.007 *	0.001 *
	(-7.29)	(-7.18)	(-2.29)	(-2.42)	(-4.88)	(-5.37)	(1.89)	(1.66)
Pdist*1000	-	-	-0.188	-0.009	-	-	-2.683 ***	-0.220 ***
			(-1.29)	(-0.56)			(-6.42)	(-5.92)
Pdist2*1000	-	-	0.000	-0.000	-	-	0.000 ***	0.000
			(0.34)	(-0.82)			(5.61)	(5.02)
Fedu	-	-	0.398 ***	0.033 ***	-	-	0.574 ***	0.037 ***
			(8.61)	(8.78)			(6.91)	(5.20)
Medu	-	-	0.214 **	-0.002	-	-	0.160	0.010
			(2.24)	(-0.30)			(1.46)	(1.09)
Ravens	-	-	0.624 ***	0.043 ***	-	-	0.289 **	0.041 ***
			(10.15)	(7.72)			(2.11)	(3.52)
Constant	7.73 ***	7.94 ***	0.440	0.336 ***			6.903 ***	0.487 ***
	(120.03)	(33.38)	(0.70)	(5.17)			(4.35)	(3.72)
R2	0.222	0.141	0.400	0.293	0.341	0.330	0.679	0.635
N	542	542	542	542	125	125	125	125
P-value (F test of excluded IVs)	-	-	0.000	0.000	-	-	0.000	0.000
P-value (overid test)	-	-	0.483				0.080	

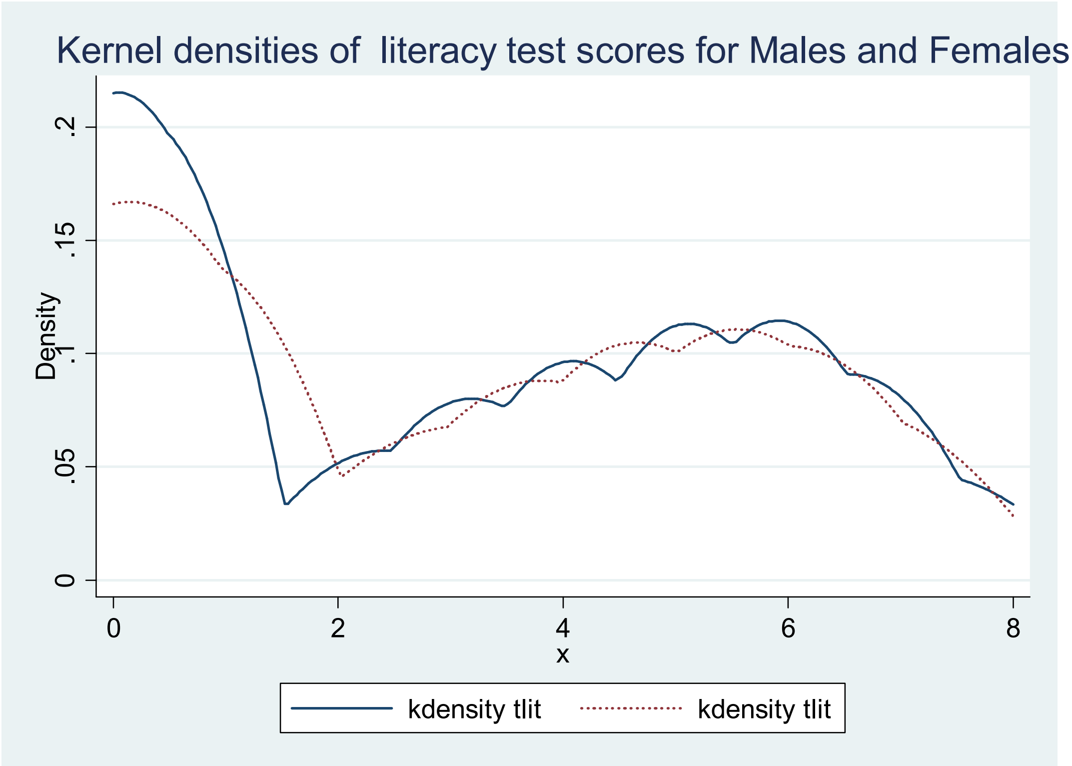
Table 13 – Household Fixed Effect Estimates of Log Earnings (All Individuals Aged 15-60)

Log earnings	**All** **OLS**		**FE**	
Male	0.564	***	0.629	***
	(4.97)		(3.73)	
Schooling	0.048	***	0.051	*
	(3.90)		(1.86)	
Exp	0.069	***	0.079	***
	(6.89)		(4.32)	
Exp2	-0.001	***	-0.002	***
	(-5.72)		(-3.49)	
Literate	-0.060		-0.055	
	(-0.43)		(-0.25)	
Constant	7.141	***	7.023	***
	(49.70)		(26.47)	
R^2	0.245		0.245	
N	401		401	
No. Of Groups	-		224	

Notes: Robust t statistics in parentheses; * significant at 10%; ** significant at 5% and *** at 1% or better;

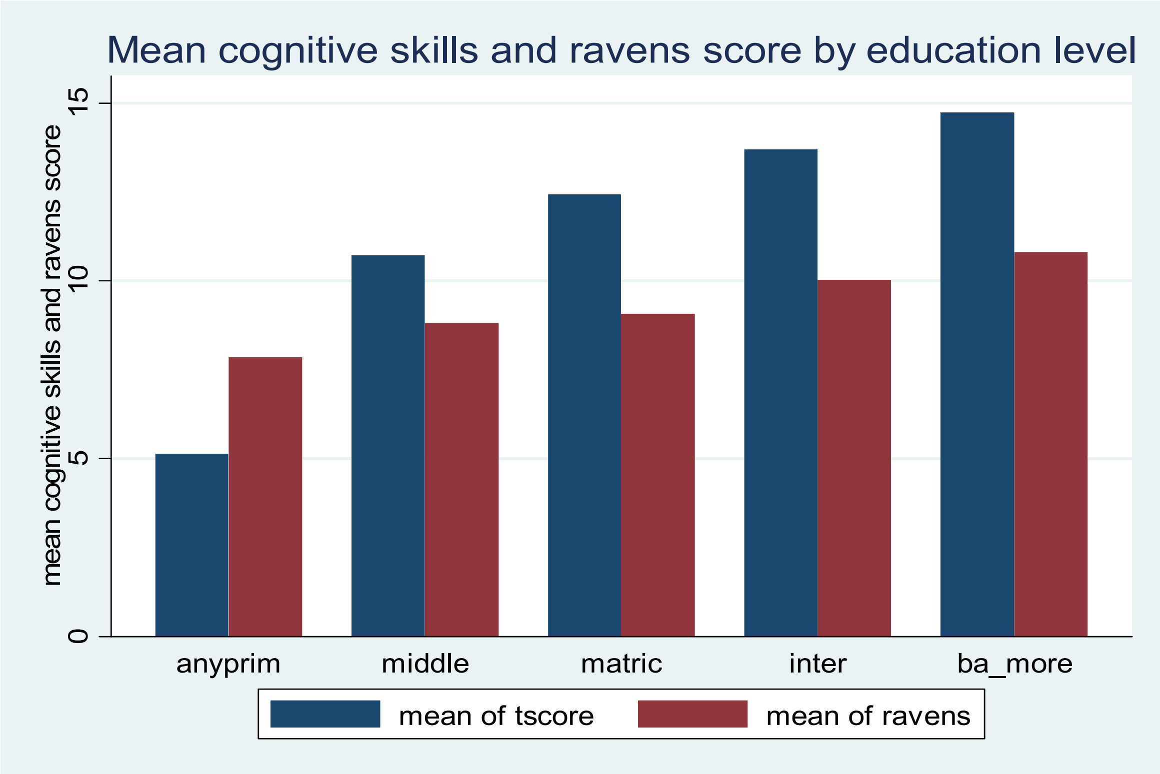
Figures

Figure 1



Note1: Solid line represents males and dotted line represents females;
Note2: Kernel densities show the distribution of a variable. Thus, a kernel density of literacy test scores shows the distribution of that score. The height of the distribution shows the frequency of individuals receiving a given score. The area under the distribution totals to 1. Thus, we can see that among males the highest frequency (compared to females) scored 0 on the test while roughly the same frequency of males and females scored between 2 and 8 points on the test.

Figure 2



Note 1: Cognitive skills score is the sum of total literacy score (max score achievable was 13) and total numeracy score (max score achievable was 13). The maximum score attainable on the ravens test was 20.

Figure 3: Males- Estimated Probability of Occupation and Education

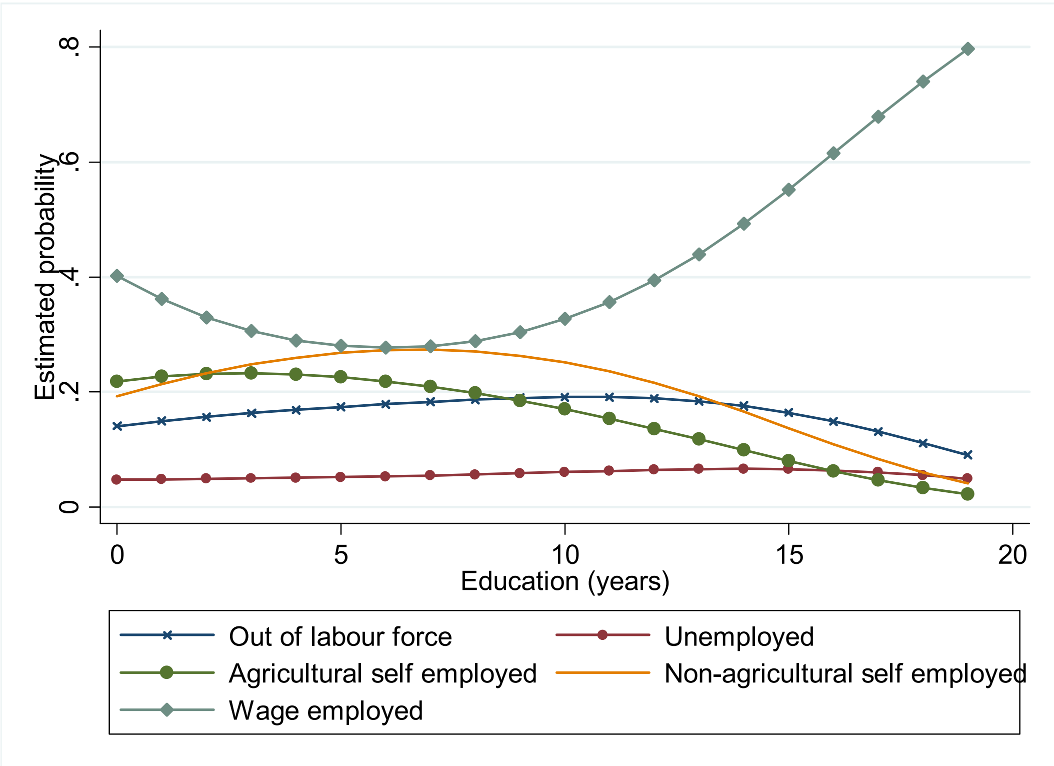
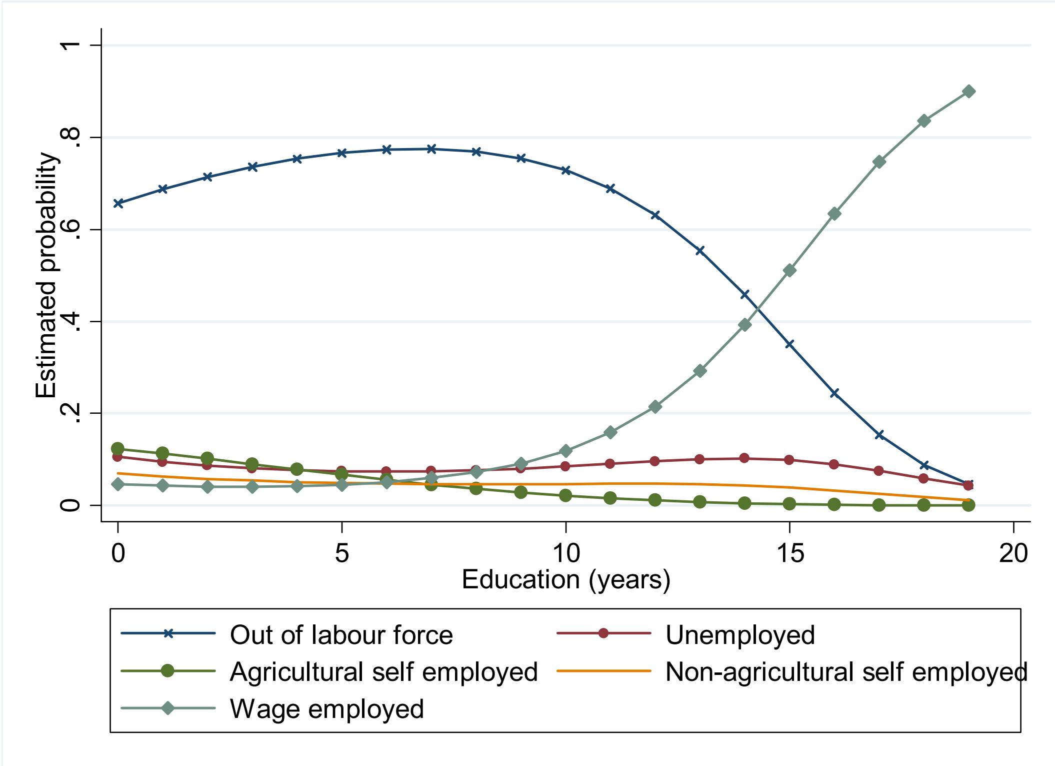


Figure 4: Females - Estimated Probability of Occupation and Education



Note: Figures 3 and 4 are based on multinomial logits reported in Table 4B.